WHAT IS CLAIMED IS:

1.   A system for compressing a data table, comprising:

a table modeller that discovers data mining models with guaranteed error bounds of at least one attribute in said data table in terms of other attributes in said data table; and

a model selector, associated with said table modeller, that selects a subset of said at least one model to form a basis upon which to compress said data table.

2.   The system as recited in Claim 1 wherein said table modeller employs classification and regression tree data mining models to model said at least one attribute.

3.   The system as recited in Claim 1 wherein said model selector employs a Bayesian network built on said at least one attribute to select relevant models for table compression.

4.   The system as recited in Claim 2 wherein construction of said models uses integrated building and pruning to exploit specified error bounds and decrease model construction time.

5.   The system as recited in Claim 1 wherein said table modeller employs a selected one of a constraint-based and a scoring-based method to generate said at least one model.

6.   The system as recited in Claim 1 wherein said model selector selects said subset based upon a compression ratio and an error bound specific for each attribute of said data table.

7.   The system as recited in Claim 1 wherein the process by which said model selector selects said subset is NP-hard.

8.   The system as recited in Claim 1 wherein said model selector selects said subset using a model built on attributes of said data table by a selected one of:

repeated calls to a maximum independent set solution algorithm, and

a greedy search algorithm.

9.   A method of compressing a data table, comprising:

discovering data mining models with guaranteed error bounds of
at least one attribute in said data table in terms of other
attributes in said data table; and

selecting a subset of said at least one model to form a basis
upon which to compress said data table.

10.   The method as recited in Claim 9 wherein said discovering
comprises employing classification and regression tree data mining
models to model said at least one attribute.

11.   The method as recited in Claim 9 wherein said discovering
comprises employing a Bayesian network built on said at least one
attribute to select relevant models for table compression.

12.   The method as recited in Claim 10 further comprising
using integrated building and pruning to exploit specified error
bounds and decrease model construction time.

13.   The method as recited in Claim 9 wherein said discovering
comprises employing a selected one of a constraint-based and a
scoring-based method to generate said at least one model.

14.   The method as recited in Claim 9 wherein said selecting
comprises selecting said subset based upon a compression ratio and
an error bound specific for each attribute of said data table.

66

15.   The method as recited in Claim 9 wherein said selecting

2   is NP-hard.

16.   The method as recited in Claim 9 wherein said selecting

2   comprises selecting said subset using a model built on attributes

3   of said data table by a selected one of:

4   repeated   calls   to   a   maximum   independent   set   solution

5   algorithm, and

6   a greedy search algorithm.

17. A database management system, comprising:

a data structure having at least one data table therein;

a database controller for allowing data to be provided to and extracted from said data structure; and

a system for compressing said at least one data table, including:

a table modeller that discovers data mining models with guaranteed error bounds of at least one attribute in said data table in terms of other attributes in said data table, and

a model selector, associated with said table modeller, that selects a subset of said at least one model to form a basis upon which to compress said data table.

18. The system as recited in Claim 17 wherein said table modeller employs classification and regression tree data mining models to model said at least one attribute.

19. The system as recited in Claim 17 wherein said model selector employs a Bayesian network built on said at least one attribute to select relevant models for table compression.

20. The system as recited in Claim 18 wherein construction of said models uses integrated building and pruning to exploit specified error bounds and decrease model construction time.

21. The system as recited in Claim 17 wherein said table

2 modeller employs a selected one of a constraint-based and a

3 scoring-based method to generate said at least one model.

22. The system as recited in Claim 17 wherein said model

2 selector selects said subset based upon a compression ratio and an

3 error bound specific for each attribute of said data table.

23. The system as recited in Claim 17 wherein the process by

which said model selector selects said subset is NP-hard.

24. The system as recited in Claim 17 wherein said model

selector selects said subset using a model built on attributes of

said data table by a selected one of:

repeated calls to a maximum independent set solution

algorithm, and

a greedy search algorithm.

69